Appendix B: Data Quality

This section addresses performance data completeness and reliability in compliance with the Office of Management and Budget's (OMB's) Circular A-11. For a fuller explanation of data limitations, data quality reviews and audits as well as improvements to data systems and collection activities, please refer to the on-line Data Quality Appendix at http://www.epa.gov/ocfo/finstatement/2006PAR (see "Supplemental Information"). This information is organized by 2006 performance measure (as presented in the *FY 2006 Performance and Accountability Report* [PAR]) and supporting database.

Data Completeness

Per OMB's definition of data completeness in its Circular A-11 (Section 230), EPA's performance data for 2006 are complete. According to OMB, performance data are complete if actual or preliminary performance is reported for every performance goal and measure. In cases where data are not currently available, OMB considers data complete if the Agency notes the year when actual performance data will be published.

For each 2006 performance target, EPA provides a measure of actual performance or a projected date when actual performance will be reported. EPA prefers not to publish preliminary data for end-of-year results where externalities could have an unpredicted impact on measured performance. As a result, in instances where a data lag exists and a date substitutes for actual data, an expectation of whether or not the annual target will be met is usually included in the goal chapter.

Limitations on Data Completeness

Output Measures versus Outcome Measures—EPA's on-going measurement improvement effort, centered in the Office of the Chief Financial Officer and in conjunction with OMB's PART process, results in the conversion of program outputs into outcome measures that track environmental results and health effects. Often, changes in environmental outcomes occur over a longer time frame than a year. Consequently, EPA emphasizes the use of performance data as a trend rather than as a 1-year result. Section II.2 (Annual Performance Goals and Measures: Detailed Results FY 2003-FY 2006) of the PAR presents these trends. In most cases where data are missing for 2006, results are reported for prior years. These trend data provide a fuller picture of Agency progress than any 1 year snapshot could capture.

Monitoring and Reporting—One reason why annual results may be missing for 2006 is because monitoring data for outcomes may be collected biennially or even less frequently. Processing the data takes additional time and results for "off-years" may be modeled. The National Emissions Inventory of Hazardous Air Pollutants, for example, is compiled every 3 years. Consequently, off-year results are projected using an emissions modeling system which accounts for economic growth and implementation of the Maximum Achievable Control Technology standards.

In the cases where performance data are collected on a calendar year basis, final results are often not published until at least the next fiscal year report. For example, data on blood-lead levels in children are collected every calendar year (by the Centers for Disease Control), but released to the public in 2-year sets. The most current data set for 2001-2002 was released in early 2005. Section II.2 (Annual Performance Goals and Measures: Detailed Results FY 2003-FY 2006), which contains more descriptive information on the performance data, indicates whether the data are collected on a fiscal or calendar year basis.

Data Reliability

Per OMB's definition of reliable data, the performance data supporting the 2006 PAR are reliable. Agency managers and decision-makers use these data on an ongoing basis in the normal course of their duties, taking into account data limitations, compensating for uncertainties, and qualifying results.

EPA has a "Quality System" in place, which encompasses formal and compulsory policies and procedures "to ensure that environmental programs and decisions are supported by the type and quality of data appropriate for their intended use and decisions involving environmental technology are supported by appropriate quality-assured engineering standards and practices." Quality system policies and documentation (e.g., Quality Management Plans), annual reviews and planning, management assessments, training, project planning, project implementation and quality assurance project plans, and verification and validation of data are all components of the Agency's Quality System. For additional information, see EPA's Quality System website at http://www.epa.gov/quality.

Because the Agency's performance data are reliable, they are not materially inadequate and, therefore, do not significantly impede the use of performance data by Agency managers.

Limitations on Data Reliability

Notwithstanding the reliability of the data presented in the FY 2006 PAR, EPA's Office of the Inspector General (OIG) and the U.S. Government Accountability Office have raised broad concerns beyond the scope of the PAR. The issues include the need to make the incorporation of data standards into data collections routine across all Agency programs, data quality associated with laboratories, and the need to be systematic in filling data gaps relating to outcome indicators presented in EPA's *DRAFT Report on the Environment.*

In addition, EPA is internally tracking three data-related management issues: data standard implementation, Permit Compliance System modernization, and Safe Drinking Water Information System Improvements. None are considered material weaknesses under the Federal Managers' Financial Integrity Act.

Section III, Management Accomplishment and Challenges, includes a discussion of issues identified as management challenges by the OIG (e.g., *Data Standards and Data Quality*) as well as the Agency's progress in addressing its self-declared management issues.

Data Standards and Data Quality—Data standards are necessary to allow EPA offices, states, tribes, and other partners to share and integrate performance information seamlessly. Without data standards, national composites can be biased, incomplete, and/or inaccurate, and development of performance outcomes can be impeded. For example, current land cleanup performance measures are based on the number of cleaned-up contaminated sites. Capturing the area or extent of land ready for use/reuse, however, would more accurately and clearly communicate the outcomes or results that EPA and its partners are striving to achieve. Some, but not all, of EPA's cleanup programs are using consistent definitions and accounting for programmatic differences in collecting placed-based information. The Agency is continuing to develop data standards and guide their implementation, for example, through an organization structured to review and approve electronic reporting systems operated by EPA and authorized state, tribal, and local government programs.

EPA and its partners are also working to ensure that data are of sufficient quality for decision making. For example, the OIG raised concerns about the integrity of results provided by laboratories' analysis of drinking water samples and the implications of poor quality data for decisions regarding human health. To address laboratory quality, EPA developed training to deter and detect improper laboratory practices. All Agency organizations, including laboratories, continue to operate under approved Quality Management Plans, which are reviewed every 3 to 4 years. For additional discussion of the Agency's efforts to address data standards and data quality, see Section III, Management Accomplishments and Challenges.

Data Gaps—The expense of collecting statistically-valid, environmental monitoring and human health data creates a challenge for the Agency to fill critical data gaps. Also, it keeps the Agency from developing important outcome measures. The Office of Water, for example, recognizes that current monitoring and assessment activities have not provided consistent and defensible national assessments of water and ecological quality (e.g., areal extent of streams, coastal waters, lakes, rivers, and wetlands impacted by nutrients, excess sedimentation, acidification, pathogens, fish and benthic animal pathologies, etc.) Collaborative efforts among EPA's Office of Research and Development and Office of Water, United States Geological Survey, National Oceanic Atmospheric Administration, and other partners will result in leveraged resources and a large-scale effort to eliminate this gap.

As part of the development of the Agency's *2006-2011 Strategic Plan*, EPA's programs were required to develop Preliminary Strategies for addressing critical data gaps, which now prevent the use of environmental outcomes. The Preliminary Strategies articulate a plan for improved environmental measures in future *Strategic Plans* as well as innovative approaches for implementation, using advanced technologies (e.g., e-reporting), collaboration and pooled resources to fill the data gaps.